LIÈGE université
Sciences
Mathématique

Contributions to spatial data analysis and Stein's method

Marie Ernst

December 17th, 2019

Introduction

Part I : Spatial dependence

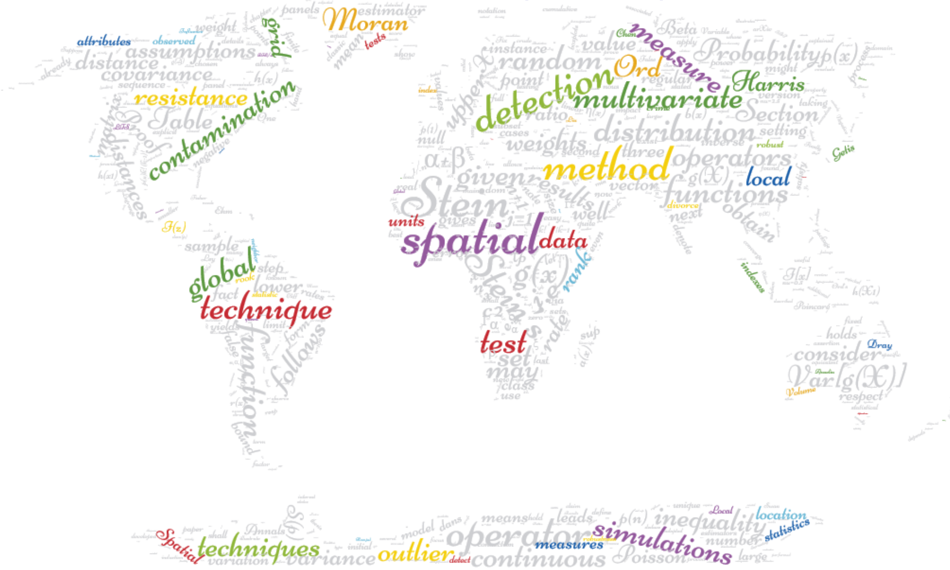
1. Outliers in spatial multivariate data
2. Robustness of tests for spatial autocorrelation

Part II: Stein's method

3. Stein differentiation
4. First order covariance identities and inequalities
5. Infinite covariance expansions
6. Stein factors and distances between distributions
7. General conclusions and perspectives

Introduction

Part 1: spatial dependence



Part 1: spatial dependence

Two research questions

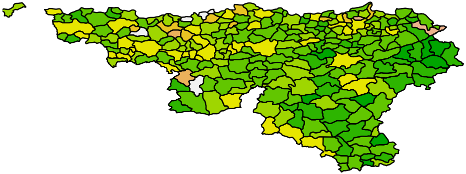
Part 1: spatial dependence

Question 1

Spatial data

Spatial data:

- geographical positions
- non spatial attributes



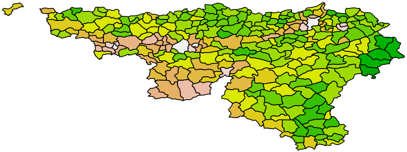
Example

Waste per capita (kg) in the Walloon region in Belgium

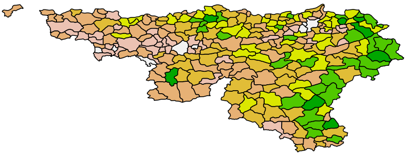
Multivariate spatial data

Example 2D

Unemployment rate



Old buildings ($\geq$ 30 years)

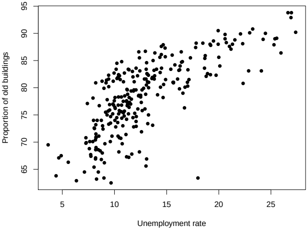


Multivariate spatial data

Example 2D

Spatial locations

Attribute representation



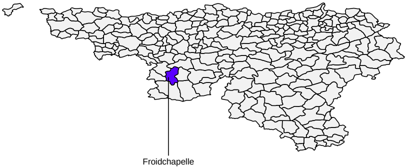
Multivariate spatial outliers

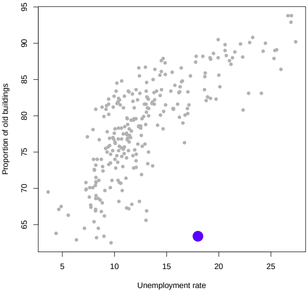
Two types of outliers (Haslett *et al.* (1991)):

- global outlier: extreme behaviour wrt all observations

Example 2D

Froidchapelle: global outlier





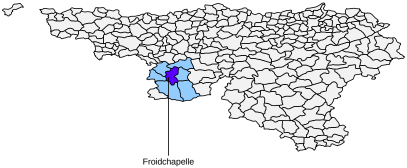
Multivariate spatial outliers

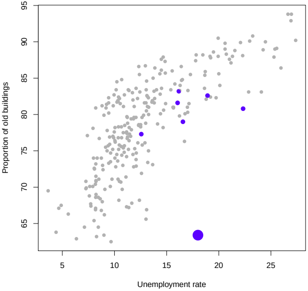
Two types of outliers (Haslett *et al.* (1991)):

- global outlier: extreme behaviour wrt all observations
- local outlier: extreme behaviour wrt its neighbours

Example 2D

Froidchapelle: global and local outlier





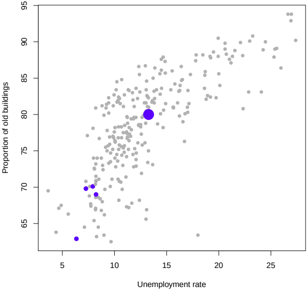
Multivariate spatial outliers

Two types of outliers (Haslett *et al.* (1991)):

- global outlier: extreme behaviour wrt all observations
- local outlier: extreme behaviour wrt its neighbours

Example 2D

Martelange: local outlier



Objectives in dimension p

Objectives in dimension p

Global outliers detection

- Geographical components not used
- Usual outlier detection techniques can be used

  $\Rightarrow$ not considered here

Objectives in dimension p

Global outliers detection

- Geographical components not used
- Usual outlier detection techniques can be used

  $\Rightarrow$ not considered here

Local outliers detection

- Review of some existing techniques
- Suggestion of an adaptation
- Comparison with examples and simulations

Review

Considered Techniques

1. Chen *et al.* (2008)
2. Harris *et al.* (2014)
3. Filzmoser *et al.* (2014)

Review

Considered Techniques

1. Chen *et al.* (2008): Componentwise median and robust Mahalanobis distance
2. Harris *et al.* (2014)
3. Filzmoser *et al.* (2014)

Review

Considered Techniques

1. Chen *et al.* (2008): Componentwise median and robust Mahalanobis distance
2. Harris *et al.* (2014): Geographically Weighted PCA with robust estimator
3. Filzmoser *et al.* (2014)

Review

Considered Techniques

1. Chen *et al.* (2008): Componentwise median and robust Mahalanobis distance
2. Harris *et al.* (2014): Geographically Weighted PCA with robust estimator
3. Filzmoser *et al.* (2014): Robust "Mahalanobis-type" detection
4. Regularized spatial detection technique: Adaptation of Filzmoser *et al.* (2014) (E. and Haesbroeck, 2017)

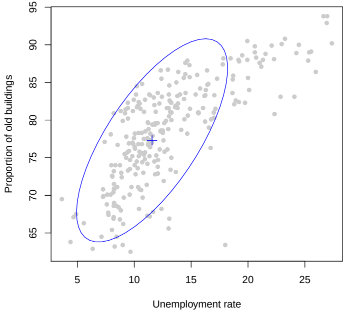
Filzmoser, Ruiz-Gazen and Thomas-Agnan (2014)

Approach

Robust "Mahalanobis-type" detection

1. Preliminary global step:
   Robust estimation of the general structure: $(\hat{\mu}, \hat{\Sigma})$

Example 2D



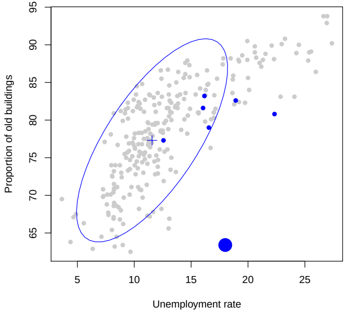
Filzmoser, Ruiz-Gazen and Thomas-Agnan (2014)

Approach

Robust "Mahalanobis-type" detection

1. Preliminary global step:
   Robust estimation of the general structure: $(\hat{\mu}, \hat{\Sigma})$
2. Local step:

Example 2D



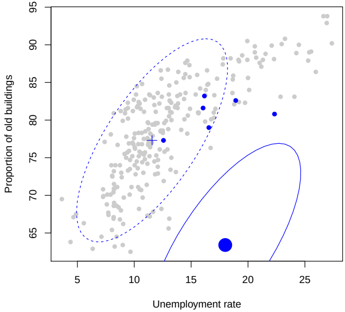
Filzmoser, Ruiz-Gazen and Thomas-Agnan (2014)

Approach

Robust "Mahalanobis-type" detection

1. Preliminary global step:
   Robust estimation of the general structure: $(\hat{\mu}, \hat{\Sigma})$
2. Local step:
   - Centring the general structure on the observation

Example 2D



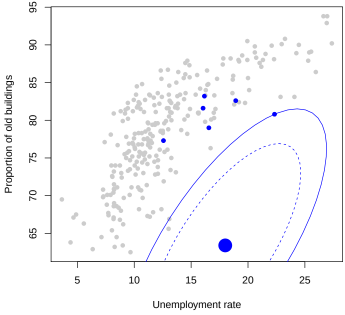
Filzmoser, Ruiz-Gazen and Thomas-Agnan (2014)

Approach

Robust "Mahalanobis-type" detection

1. Preliminary global step:
   Robust estimation of the general structure: $(\hat{\mu}, \hat{\Sigma})$
2. Local step:
   - Centring the general structure on the observation
   - Determination of the ellipsoid containing the *next neighbour*

Example 2D



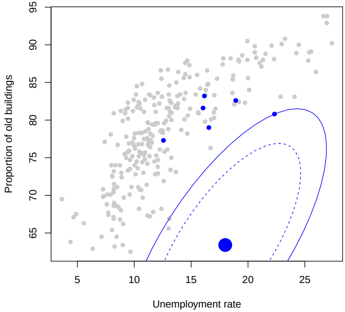
Filzmoser, Ruiz-Gazen and Thomas-Agnan (2014)

Approach

Robust "Mahalanobis-type" detection

1. Preliminary global step:
   Robust estimation of the general structure: $(\hat{\mu}, \hat{\Sigma})$
2. Local step:
   - Centring the general structure on the observation
   - Determination of the ellipsoid containing the *next neighbour*
   - If its tolerance level is larger than a theoretical quantile
     $\Rightarrow$ **local outlier**

Example 2D



Regularized spatial detection technique

(E. and Haesbroeck, 2017)

Approach: adaptation of Filzmoser *et al.* (2014)

Work with local structure and only on the most homogeneous neighbourhoods

Regularized spatial detection technique

(E. and Haesbroeck, 2017)

Approach: adaptation of Filzmoser *et al.* (2014)

Work with local structure and only on the most homogeneous neighbourhoods

1. Local step:
   - Estimation of the local structure: $(\hat{\mu}_i, \hat{\Sigma}_i)$ with robust and regularized estimators

Regularized spatial detection technique

(E. and Haesbroeck, 2017)

Approach: adaptation of Filzmoser *et al.* (2014)

Work with local structure and only on the most homogeneous neighbourhoods

1. Local step:
   - Estimation of the local structure: $(\hat{\mu}_i, \hat{\Sigma}_i)$ with robust and regularized estimators
   - Homogeneity measure: $\det(\hat{\Sigma}_i)$

Regularized spatial detection technique

(E. and Haesbroeck, 2017)

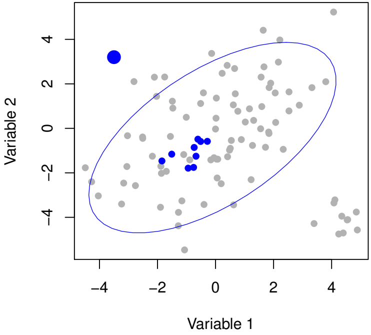
Approach: adaptation of Filzmoser *et al.* (2014)

Work with local structure and only on the most homogeneous neighbourhoods

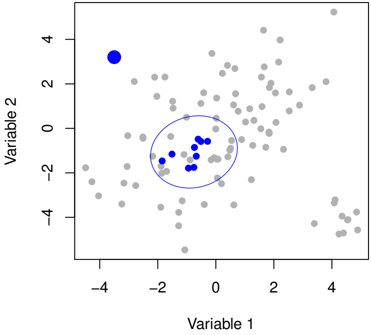
1. Local step:
   - Estimation of the local structure: $(\hat{\mu}_i, \hat{\Sigma}_i)$ with robust and regularized estimators
   - Homogeneity measure: $\det(\hat{\Sigma}_i)$
2. Global step: Selection of 10%, 20%, . . . of smallest values
3. Local step: work only on *selected* neighbourhoods
   - Centring the *local* structure on the observation
   - Determination of the ellipsoid containing the next neighbour
   - If its tolerance level is larger than an empirical quantile $\Rightarrow$ **local outlier**

Illustration

Outliers for Filzmoser et al.

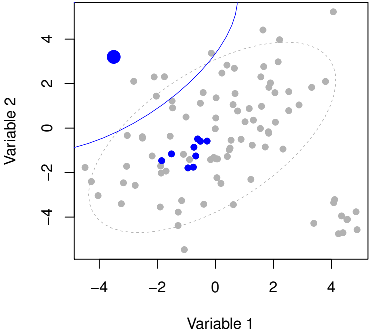


Outliers for regularization
Test on 10% of neighbourhoods



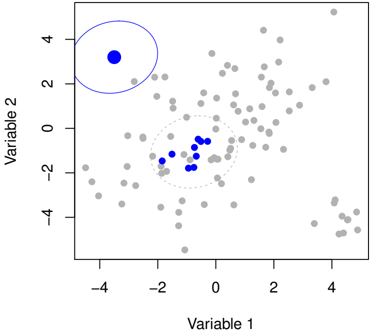
Illustration

Outliers for Filzmoser et al.



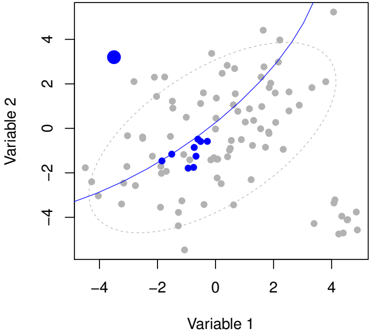
Outliers for regularization

Test on 10% of neighbourhoods

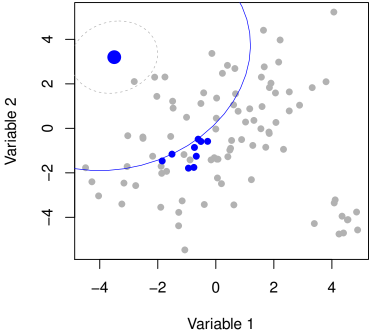


Illustration

Outliers for Filzmoser et al.

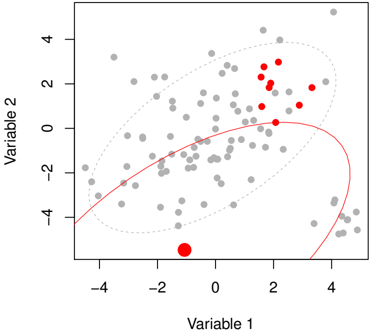


Outliers for regularization
Test on 10% of neighbourhoods

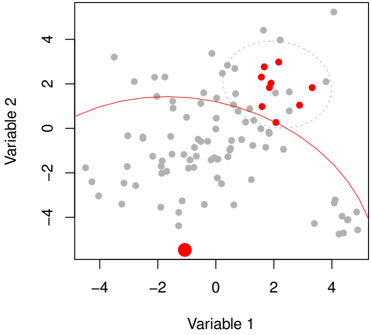


Illustration

Outliers for Filzmoser et al.

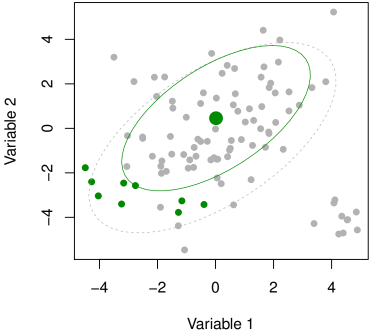


Outliers for regularization
Test on 20% of neighbourhoods



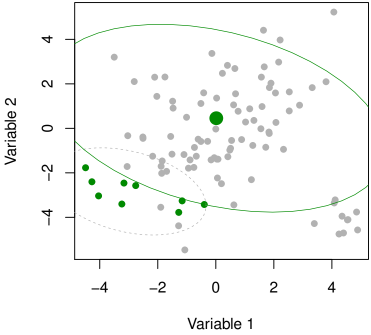
Illustration

Outliers for Filzmoser et al.



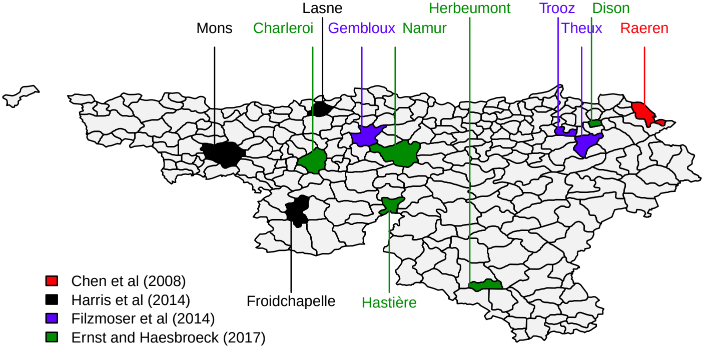
Outliers for regularization

Test on 30% of neighbourhoods



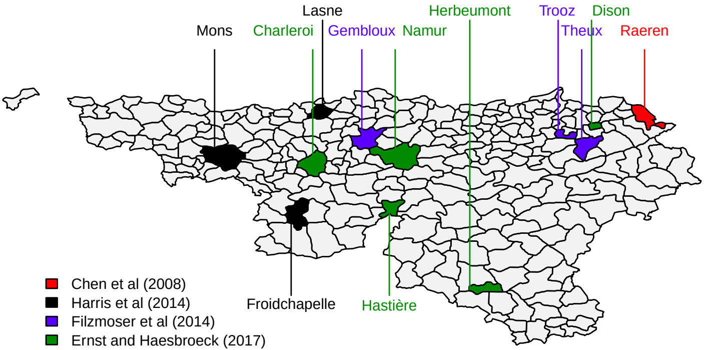
Illustrations

Wallonia: 14 socio-economic variables for the 262 municipalities



Illustrations

Wallonia: 14 socio-economic variables for the 262 municipalities



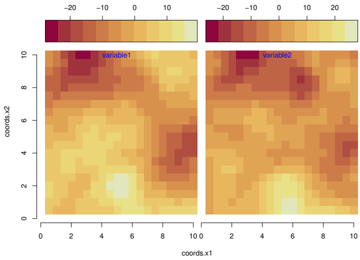
⇝ Simulations for the comparison

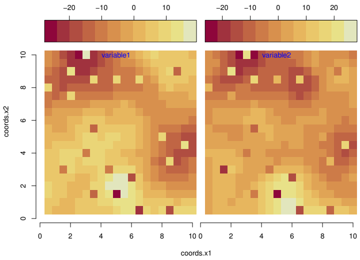
Simulations

Generate spatial data of p variables for n locations (grid or Walloon municipalities)

Simulation set-up

- Matérn model to generate spatial data varying the overall smoothness
- Contamination by swapping observations with high/small PCA scores[1]





[1] Harris *et al.* (2014)

Results

- Harris et al. (2014) wrongly flags too many good observations as local outliers.
- Chen et al. (2008) handles well the regular domain with the less smooth design.
- Filzmoser et al (2014) outperforms the two previous techniques in most configurations (smoother variable and/or irregular domain).
- The adaptation has similar results as the initial technique; these results being dependent on the homogeneity constraint that we set.

Spatial autocorrelation

Question 2: When do we have to consider spatial techniques?

Spatial autocorrelation

Question 2: When do we have to consider spatial techniques?

When the i.i.d. assumption is no longer satisfied?

Are the values randomly assigned to locations?

Spatial autocorrelation

Question 2: When do we have to consider spatial techniques?

When the i.i.d. assumption is no longer satisfied?

Are the values randomly assigned to locations?

$\rightsquigarrow$ check spatial autocorrelation

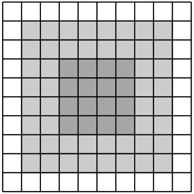
Spatial autocorrelation

Question 2: When do we have to consider spatial techniques?

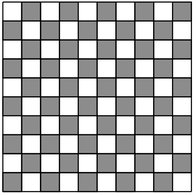
When the i.i.d. assumption is no longer satisfied?

Are the values randomly assigned to locations?

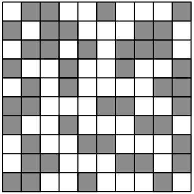
$\rightsquigarrow$ check spatial autocorrelation



Positive spatial autocorrelation



Negative spatial autocorrelation



No spatial autocorrelation

Neighbours

Weighting matrix W

Locations s_i and s_j are neighbours if and only if $w_{ij} > 0$.
Otherwise, $w_{ij} = 0$.

Neighbours

Weighting matrix W

Locations s_i and s_j are neighbours if and only if $w_{ij} > 0$.
Otherwise, $w_{ij} = 0$.

Different choices:

- Binary weights,
- Row-standardized,
- Globally standardized, . . .

Neighbours

Weighting matrix W

Locations s_i and s_j are neighbours if and only if $w_{ij} > 0$.
Otherwise, $w_{ij} = 0$.

Different choices:

- Binary weights,
- Row-standardized,
- Globally standardized, ...

Convention: zero diagonal and $S_0 = \sum_i \sum_j w_{ij}$.

Measures of spatial autocorrelation

Sample data points $\boldsymbol{z} = \{z_1, \ldots, z_n\}$ observed at spatial locations $\{s_1, \ldots, s_n\}$

Measures of spatial autocorrelation

Sample data points $\boldsymbol{z} = \{z_1, \ldots, z_n\}$ observed at spatial locations $\{s_1, \ldots, s_n\}$

Moran's Index (1950)

$$I(\boldsymbol{z}) = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(z_i - \bar{z})(z_j - \bar{z})}{\sum_{i=1}^{n}(z_i - \bar{z})^2}$$

Measures of spatial autocorrelation

Sample data points $\boldsymbol{z} = \{z_1, \ldots, z_n\}$ observed at spatial locations $\{s_1, \ldots, s_n\}$

Moran's Index (1950)

$$I(\boldsymbol{z}) = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (z_i - \bar{z})(z_j - \bar{z})}{\sum_{i=1}^{n} (z_i - \bar{z})^2}$$

Measures of spatial autocorrelation

Sample data points $\boldsymbol{z} = \{z_1, \ldots, z_n\}$ observed at spatial locations $\{s_1, \ldots, s_n\}$

Moran's Index (1950)

$$I(\boldsymbol{z}) = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (z_i - \bar{z})(z_j - \bar{z})}{\sum_{i=1}^{n} (z_i - \bar{z})^2}$$

Geary's ratio (1954)

$$c(\boldsymbol{z}) = \frac{n-1}{2S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (z_i - z_j)^2}{\sum_{i=1}^{n} (z_i - \bar{z})^2}$$

Getis and Ord's statistics (1992)

$$G(\boldsymbol{z}) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} z_i z_j}{\sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} z_i z_j}$$

Inference

Tests based on asymptotic normality

Without spatial autocorrelation, I, c and G are **asymptotically Gaussian** under normality (N) and/or randomisation (R) assumption.

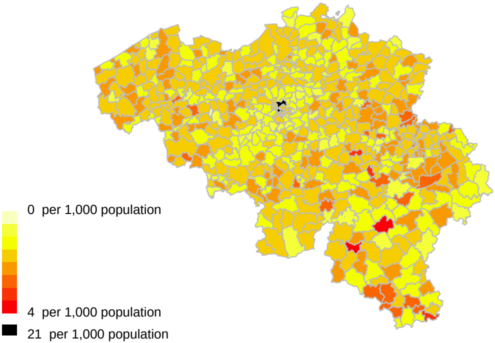
Tests based on asymptotic normality

Without spatial autocorrelation, I, c and G are **asymptotically Gaussian** under normality (N) and/or randomisation (R) assumption.

Moran's I	**Geary's c**	**Getis and Ord's G**
Test under R	Test under R	Test under R
Test under N	Test under N	
Permutation test	Permutation test	Permutation test
Dray's test		

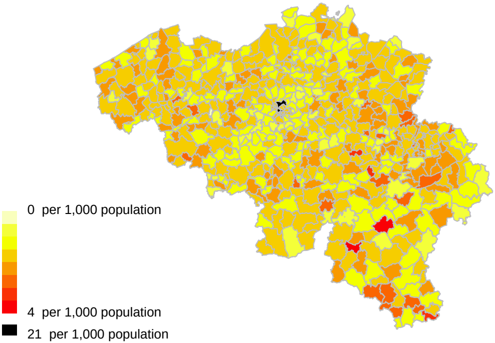
Robustness: example

Crude divorce rate in Belgium



Robustness: example

Crude divorce rate in Belgium



- Moran's $I = 0.01$
- $E_{H_0}[I] = -0.0017$
- Range values: $-0.67 \leq I \leq 1.16$
- p-value=0.25 (0.32) under R (under N)

Robustness of the tests

Robustness (Huber, 1981)

Insensitivity to small deviations from assumptions and more precisely, outlier resistance

Classic robustness tools

- Breakdown point (Hampel 1971)
- Influence function (Hampel et al. 1986)

Characteristic

Based on functionals ($n \to \infty$)

Robustness of the tests

Robustness (Huber, 1981)

Insensitivity to small deviations from assumptions and more precisely, outlier resistance

Classic robustness tools

- Breakdown point (Hampel 1971)
- Influence function (Hampel et al. 1986)

Characteristic

Based on functionals ($n \to \infty$) $\rightsquigarrow$ favour empirical tools (finite n)

Empirical influence function of the p-value

(Lambert 1981)

Definition

$$EIF(\xi, i) = \frac{\text{p-value}(\mathbf{z} + \xi \mathbf{e_i}) - \text{p-value}(\mathbf{z})}{1/n}$$

Empirical influence function of the p-value

(Lambert 1981)

Definition

$$EIF(\xi, i) = \frac{\text{p-value}(\mathbf{z} + \xi \mathbf{e_i}) - \text{p-value}(\mathbf{z})}{1/n}$$

Proposition (Chapter 2, Prop. 2.4.1)

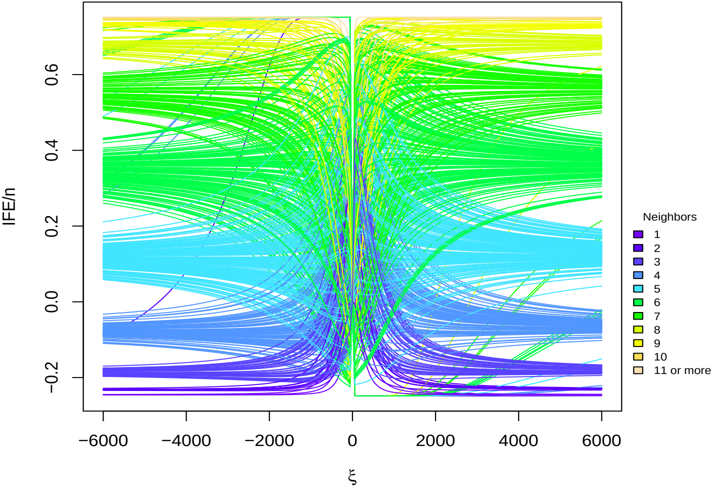
The EIF of the p-value of unilateral tests based on asymptotic normality for Moran's index is explicitly given by

$$\begin{aligned} EIF(\xi, i; I) &= n \left[\Phi\left(-\frac{I(\mathbf{z} + \xi e_i) - \mathbb{E}[I]}{\sigma[I(\mathbf{z} + \xi e_i)]} \right) - \Phi\left(-\frac{I(\mathbf{z}) - \mathbb{E}[I]}{\sigma[I(\mathbf{z})]} \right) \right] \\ &\to n \left[\Phi\left(\frac{2\,(n w_{i\bullet} - S_0)}{(n-1) S_0 \sigma[I]} \right) - \Phi\left(-\frac{I(\mathbf{z}) - \mathbb{E}[I]}{\sigma[I]} \right) \right] \end{aligned}$$

as ξ tends to infinity.

Robustness: example

Crude divorce rate: hair-plot (Genton and Ruiz-Gazen, 2010)



Resistance of a test

(Ylvisaker 1977)

Definition

Resistance to acceptance (resp. rejection): smallest proportion of the data that must be corrupted to guarantee the acceptance (rejection) of H_0.

Resistance of a test

(Ylvisaker 1977)

Definition

Resistance to acceptance (resp. rejection): smallest proportion of the data that must be corrupted to guarantee the acceptance (rejection) of H_0.

Proposition (Chapter 2, Prop. 2.4.2 - 2.4.3)

- Resistance to acceptance is $1/n$ for both asymptotic tests
- The resistance to rejection is m/n where m is the size of the smallest subset $A \subseteq \{1, \ldots, n\}$ which satisfies

$$\begin{cases} \frac{n^2 w_A - 2nm(w_A + w_B) + m^2 S_0}{S_0 m(n-m)} > \frac{-1}{n-1} + \sigma_N[I] z_{1-\alpha} & \text{under N} \\ \frac{n^2 w_A - 2mn(w_A + w_B) + m^2 S_0}{S_0 m(n-m)} > \frac{-1}{n-1} + \sqrt{a_1 - a_2 \frac{n^2 - 3nm + 3m}{(n-m)m}} z_{1-\alpha} & \text{under R} \end{cases}$$

Robustness: example

Example: Belgium

- *Resistance to acceptance*: for divorces, the contaminated rate of Brussels modified the result into an acceptance of H_0.
  If the "true" value is associated with Brussels[2], Moran's I is 0.14 (p-value < 0.0001) instead of 0.01 (p-value ≥ 0.25).

[2] Brussels: 24 local divorces vs 3698 divorces

Robustness: example

Example: Belgium

- *Resistance to acceptance*: for divorces, the contaminated rate of Brussels modified the result into an acceptance of H_0. If the "true" value is associated with Brussels[2], Moran's I is 0.14 (p-value < 0.0001) instead of 0.01 (p-value ≥ 0.25).
- *Resistance to rejection*: two corrupted neighbours are enough to always reject H_0.

[2] Brussels: 24 local divorces vs 3698 divorces

Robustness: example

Example: Belgium

- *Resistance to acceptance*: for divorces, the contaminated rate of Brussels modified the result into an acceptance of H_0.
  If the "true" value is associated with Brussels[2], Moran's I is 0.14 (p-value < 0.0001) instead of 0.01 (p-value ≥ 0.25).
- *Resistance to rejection*: two corrupted neighbours are enough to always reject H_0.

Other tests

The lack of robustness of the other tests is similarly proved.

[2] Brussels: 24 local divorces vs 3698 divorces

Robust alternative 1

Rank Moran index I_r

Idea: replace observations by their rank

Robust alternative 1

Rank Moran index I_r

Idea: replace observations by their rank

Proposition (Chapter 2, Prop 2.5.2)

$$EIF(\xi, i; I_r) = n \left[\Phi\left(\frac{I_r(\mathbf{z}) - E[I_r]}{\sigma[I_r]} \right) - \Phi\left(\frac{I_r(\mathbf{z}) - E[I_r]}{\sigma[I_r]} + \frac{e}{\sigma[I_r]} \right) \right]$$

where $e = Ir(\mathbf{z} + \xi e_i) - Ir(\mathbf{z})$ is explicit.
The impact of contamination on the p-value is limited.

Robust alternative 1

Rank Moran index I_r

Idea: replace observations by their rank

Proposition (Chapter 2, Prop 2.5.2)

$$EIF(\xi, i; I_r) = n\left[\Phi\left(\frac{I_r(\mathbf{z}) - E[I_r]}{\sigma[I_r]}\right) - \Phi\left(\frac{I_r(\mathbf{z}) - E[I_r]}{\sigma[I_r]} + \frac{e}{\sigma[I_r]}\right)\right]$$

where $e = Ir(\mathbf{z} + \xi e_i) - Ir(\mathbf{z})$ is explicit.
The impact of contamination on the p-value is limited.

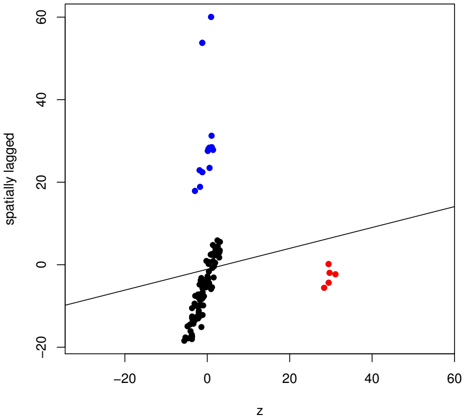
Estimated resistances (Belgium)

- Resistance to rejection: at most $14/589$
- Resistance to acceptance: $6/589$

Robust alternative 2

Moran scatterplot

Moran can be interpreted as the slope in a OLS regression of spatially lagged observations over **z**.

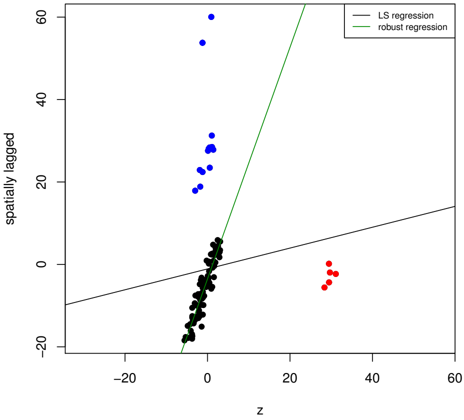


Robust alternative 2

Moran scatterplot

Moran can be interpreted as the slope in a OLS regression of spatially lagged observations over **z**.

⇝ **Idea**: use robust regression



Robust and efficient regression estimation

Different methods

- S-estimator (Rousseew and Yohai, 1984)
- Least Trimmed Squares (LTS - Rousseeuw, 1985)
- MM-estimator (Yohai, 1987)
- Robust and Efficient Weighted Least Squares Estimator (REWLSE - Gervini and Yohai, 2002)

Robust and efficient regression estimation

Different methods

- S-estimator (Rousseew and Yohai, 1984)
- Least Trimmed Squares (LTS - Rousseeuw, 1985)
- MM-estimator (Yohai, 1987)
- Robust and Efficient Weighted Least Squares Estimator (REWLSE - Gervini and Yohai, 2002)

Robustness of adapted Moran index

Due to robust properties of regression estimators,

- the impact on the p-value is zero almost everywhere;
- the resistance is m/n where m is the minimal number of observations for which the joint neighbourhoods contains at most 50% of the points in Moran scatterplot.

Simulation study

Efficiency of robust tests

Comparison of level and power of robust and classic tests.

Simulation study

Efficiency of robust tests

Comparison of level and power of robust and classic tests.

Spatial autoregressive model

$$Z = \rho WZ + \varepsilon$$

where ρ is the spatial correlation coefficient.

Simulation study

Efficiency of robust tests

Comparison of level and power of robust and classic tests.

Spatial autoregressive model

$$Z = \rho W Z + \varepsilon$$

where ρ is the spatial correlation coefficient.

Results

- Comparable power and level for all tests
- Power increases with n
- Power increases with ρ

End of Part 1


Moran
attributes
observed
assumptions
distance
covariance
resistance
contamination
grid
Table
global
technique
function
follows
detection
multivariate
measure
Ord
Harris
Probability
random
distribution
method
operators
local
functions
spatial
data
units
Stein
rank
test
Section
Spatial
techniques
outlier
operator
continuous
measures
simulations
inequality
location
statistics
variance

Part 2: Stein's method

Gaussian case

Stein's identity

If $X \sim \mathcal{N}(0,1)$, then

$$\mathbb{E}[Xg(X)] = \mathbb{E}[g'(X)] \ \forall g \text{ s.t. } E[|g'(X)|] < \infty$$

Gaussian case

Stein's identity

If $X \sim \mathcal{N}(0,1)$, then

$$\mathbb{E}[Xg(X)] = \mathbb{E}[g'(X)] \; \forall g \text{ s.t. } E[|g'(X)|] < \infty$$

Stein operator

The associated *Stein operator* is defined by

$$\mathcal{T}g(x) = xg(x) - g'(x)$$

and $\mathbb{E}[\mathcal{T}g(X)] = 0$ for any appropriate g.

Let $X \sim p$. The derivative-type operators are

$$\Delta^{\ell} f(x) = \begin{cases} f'(x) & \text{if } \ell = 0; \\ f(x+1) - f(x) & \text{if } \ell = +1; \\ f(x) - f(x-1) & \text{if } \ell = -1. \end{cases}$$

Formalism

Let $X \sim p$. The derivative-type operators are

$$\Delta^{\ell} f(x) = \begin{cases} f'(x) & \text{if } \ell = 0; \\ f(x+1) - f(x) & \text{if } \ell = +1; \\ f(x) - f(x-1) & \text{if } \ell = -1. \end{cases}$$

Canonical Stein operator

For any $f : \mathbb{R} \to \mathbb{R}$, the *canonical Stein operator* is defined as

$$\mathcal{T}_p^{\ell} : f(x) \mapsto \mathcal{T}_p^{\ell} f(x) := \frac{\Delta^{\ell}(f(x)p(x))}{p(x)}.$$

For any f from the *canonical Stein class*, $\mathbb{E}[\mathcal{T}_p f(X)] = 0$.

Formalism

Let $X \sim p$. The derivative-type operators are

$$\Delta^{\ell} f(x) = \begin{cases} f'(x) & \text{if } \ell = 0; \\ f(x+1) - f(x) & \text{if } \ell = +1; \\ f(x) - f(x-1) & \text{if } \ell = -1. \end{cases}$$

Canonical Stein operator

For any $f : \mathbb{R} \to \mathbb{R}$, the *canonical Stein operator* is defined as

$$\mathcal{T}_p^{\ell} : f(x) \mapsto \mathcal{T}_p^{\ell} f(x) := \frac{\Delta^{\ell}(f(x)p(x))}{p(x)}.$$

For any f from the *canonical Stein class*, $\mathbb{E}[\mathcal{T}_p f(X)] = 0$.

Particular case: the score function

$$\rho_p^{\ell}(x) = \mathcal{T}_p^{\ell} 1(x) = \frac{\Delta^{\ell} p(x)}{p(x)}$$

Example: Poisson density

$p(x) = e^{-\lambda}\lambda^x/x!$ for $x \in \mathbb{N}$ and $\ell = \pm 1$

Example: Poisson density

$p(x) = e^{-\lambda}\lambda^x/x!$ for $x \in \mathbb{N}$ and $\ell = \pm 1$

Canonical Stein operators

- $\mathcal{T}_p^+ f(x) = \dfrac{\Delta^+(f(x)p(x))}{p(x)} = f(x+1)\dfrac{\lambda}{x+1} - f(x)$
- $\mathcal{T}_p^- f(x) = \dfrac{\Delta^-(f(x)p(x))}{p(x)} = f(x) - f(x-1)\dfrac{x}{\lambda}$

Example: Poisson density

$p(x) = e^{-\lambda}\lambda^x/x!$ for $x \in \mathbb{N}$ and $\ell = \pm 1$

Canonical Stein operators

- $\mathcal{T}_p^+ f(x) = \dfrac{\Delta^+(f(x)p(x))}{p(x)} \quad = f(x+1)\dfrac{\lambda}{x+1} - f(x)$
- $\mathcal{T}_p^- f(x) = \dfrac{\Delta^-(f(x)p(x))}{p(x)} = f(x) - f(x-1)\dfrac{x}{\lambda}$

The score functions

- $\rho_p^+(x) = \mathcal{T}_p^+ 1(x) = \dfrac{\lambda}{x+1} - 1$
- $\rho_p^- f(x) = \mathcal{T}_p^- 1(x) = 1 - \dfrac{x}{\lambda}$

Pseudo inverse Stein operator

The *canonical pseudo inverse Stein operator* for the operator $\mathcal{T}_p^\ell$ is

$$\mathcal{L}_p^\ell : h(x) \mapsto \mathcal{L}_p^\ell h(x) := \frac{1}{p(x)} \int_a^{x-a_\ell} (h(u) - \mathbb{E}[h(X)]) p(u) \mu(du)$$

where $a_\ell = \mathbb{I}[\ell = 1]$.

Properties

- $\mathcal{L}_p^\ell \mathcal{T}_p^\ell f(x) = f(x)$
- $\mathcal{T}_p^\ell \mathcal{L}_p^\ell h(x) = h(x) - \mathbb{E}[h(X)]$

Pseudo inverse Stein operator

The *canonical pseudo inverse Stein operator* for the operator $\mathcal{T}_p^\ell$ is

$$\mathcal{L}_p^\ell : h(x) \mapsto \mathcal{L}_p^\ell h(x) := \frac{1}{p(x)} \int_a^{x-a_\ell} (h(u) - \mathbb{E}[h(X)])p(u)\mu(du)$$

where $a_\ell = \mathbb{I}[\ell = 1]$.

Properties

- $\mathcal{L}_p^\ell \mathcal{T}_p^\ell f(x) = f(x)$
- $\mathcal{T}_p^\ell \mathcal{L}_p^\ell h(x) = h(x) - \mathbb{E}[h(X)]$

Particular case: the Stein kernel

$$\tau_p^\ell(x) = -\mathcal{L}_p^\ell \mathrm{Id}(x)$$

Pseudo inverse Stein operator

The *canonical pseudo inverse Stein operator* for the operator $\mathcal{T}_p^\ell$ is

$$\mathcal{L}_p^\ell : h(x) \mapsto \mathcal{L}_p^\ell h(x) := \frac{1}{p(x)} \int_a^{x-a_\ell} (h(u) - \mathbb{E}[h(X)]) p(u) \mu(du)$$

where $a_\ell = \mathbb{I}[\ell = 1]$.

Properties

- $\mathcal{L}_p^\ell \mathcal{T}_p^\ell f(x) = f(x)$
- $\mathcal{T}_p^\ell \mathcal{L}_p^\ell h(x) = h(x) - \mathbb{E}[h(X)]$

Particular case: the Stein kernel

$$\tau_p^\ell(x) = -\mathcal{L}_p^\ell \mathrm{Id}(x)$$

Poisson density

- $\tau_p^+(x) = x$ and $\tau_p^-(x) = \lambda$

Stein equation

Standardized Stein operator

$$\mathcal{A}g(x) = \mathcal{T}_p^{\ell}\left(f(\cdot)g(\cdot - \ell)\right)(x) = \mathcal{T}_p^{\ell}f(x)g(x) + f(x)\Delta^{-\ell}g(x)$$

Stein equation

Standardized Stein operator

$$\mathcal{A}g(x) = \mathcal{T}_p^\ell \left(f(\cdot)g(\cdot - \ell)\right)(x) = \mathcal{T}_p^\ell f(x)g(x) + f(x)\Delta^{-\ell}g(x)$$

Stein covariance identity (IBP 1)

For all "appropriate" f and g,

$$\mathbb{E}\left[\left(\mathcal{T}_p^\ell f(X)\right)g(X)\right] = -\mathbb{E}\left[f(X)\Delta^{-\ell}g(X)\right]$$

Stein equation

Standardized Stein operator

$$\mathcal{A}g(x) = \mathcal{T}_p^{\ell}\left(f(\cdot)g(\cdot - \ell)\right)(x) = \mathcal{T}_p^{\ell} f(x)g(x) + f(x)\Delta^{-\ell}g(x)$$

Stein covariance identity (IBP 1)

For all "appropriate" f and g,

$$\mathbb{E}\left[\left(\mathcal{T}_p^{\ell} f(X)\right) g(X)\right] = -\mathbb{E}\left[f(X)\Delta^{-\ell}g(X)\right]$$

Stein covariance identity (IBP 2)

For all "appropriate" f and g,

$$\mathrm{Cov}\left[f(X), g(X)\right] = \mathbb{E}\left[-\left(\mathcal{L}_p^{\ell} f(X)\right)\Delta^{-\ell}g(X)\right]$$

Representations of the inverse operator

E., Reinert and Swan (2019)

- $$\mathcal{L}_p^\ell f(x) = \mathbb{E}\left[(f(X) - \mathbb{E}[f(X)])\frac{\mathbb{I}[X \leq x - a_\ell]}{p(x)}\right] \quad \text{(Definition)}$$

Representations of the inverse operator

E., Reinert and Swan (2019)

- $$\mathcal{L}_p^\ell f(x) = \mathbb{E}\left[(f(X) - \mathbb{E}[f(X)])\frac{\mathbb{I}[X \leq x - a_\ell]}{p(x)}\right] \quad \text{(Definition)}$$
- $$-\mathcal{L}_p^\ell f(x) = \mathbb{E}\left[(f(X_2) - f(X_1))\Phi(X_1, x, X_2)\right] \quad \text{(Rep. I)}$$

where $\Phi(u, x, v) = \mathbb{I}[u + a_\ell \leq x \leq v - b_\ell]/p(x)$

Representations of the inverse operator

E., Reinert and Swan (2019)

- $\mathcal{L}_p^\ell f(x) = \mathbb{E}\left[(f(X) - \mathbb{E}[f(X)])\frac{\mathbb{I}[X \leq x - a_\ell]}{p(x)}\right]$ (Definition)
- $-\mathcal{L}_p^\ell f(x) = \mathbb{E}\left[(f(X_2) - f(X_1))\Phi(X_1, x, X_2)\right]$ (Rep. I)

where $\Phi(u, x, v) = \mathbb{I}[u + a_\ell \leq x \leq v - b_\ell]/p(x)$

- $-\mathcal{L}_p^\ell f(x) = \mathbb{E}\left[\frac{K_p^\ell(X, x)}{p(X)p(x)}\Delta^{-\ell} f(X)\right]$ (Rep. II)

where $K_p^\ell(x, x') = \mathbb{P}[X \leq (x \wedge x') - a_\ell]\mathbb{P}[X \geq (x \vee x') + b_\ell]$

Representations of the inverse operator

E., Reinert and Swan (2019)

- $$\mathcal{L}_p^\ell f(x) = \mathbb{E}\left[(f(X) - \mathbb{E}[f(X)])\frac{\mathbb{I}[X \le x - a_\ell]}{p(x)}\right] \quad \text{(Definition)}$$
- $$-\mathcal{L}_p^\ell f(x) = \mathbb{E}\left[(f(X_2) - f(X_1))\Phi(X_1, x, X_2)\right] \quad \text{(Rep. I)}$$

where $\Phi(u, x, v) = \mathbb{I}[u + a_\ell \le x \le v - b_\ell]/p(x)$

- $$-\mathcal{L}_p^\ell f(x) = \mathbb{E}\left[\frac{K_p^\ell(X, x)}{p(X)p(x)}\Delta^{-\ell} f(X)\right] \quad \text{(Rep. II)}$$

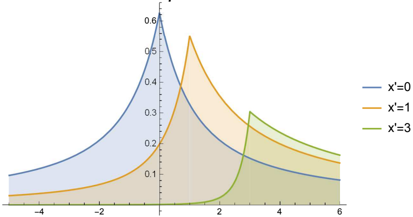
where $K_p^\ell(x, x') = \mathbb{P}[X \le (x \wedge x') - a_\ell]\mathbb{P}[X \ge (x \vee x') + b_\ell]$

References for $\ell = 0$:

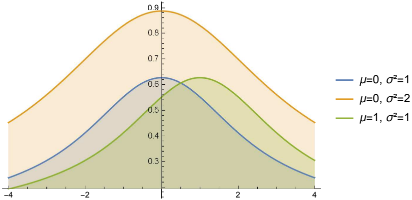
- Rep. I is non-explicitly given in Chatterjee and Shao (2011)
- Rep. II is available in Saumard (2019)
- Symmetric kernel K_p^0: first appearance attributed to Höffding (1940)

Examples

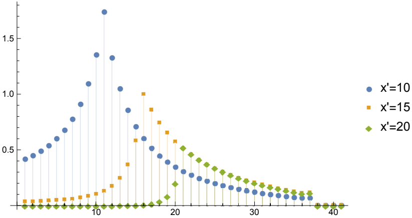
$x \mapsto K_p^\ell(x, x')/p(x)$ $\qquad$ $x \mapsto K_p^\ell(x, x)/p(x)$



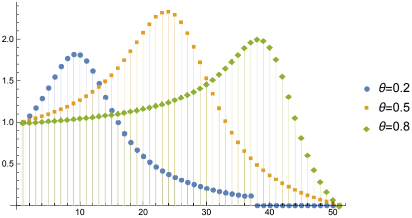
$\mathcal{N}(0, 1)$



$\mathcal{N}(\mu, \sigma^2)$



$\mathrm{Bin}(50, 0.2)$



$\mathrm{Bin}(50, \theta)$

Applications

Application 1: First order variance bounds

Family of first order lower and upper covariance bounds for functionals of arbitrary univariate distributions.

$\rightsquigarrow$ E., Reinert and Swan (2019a), Bernoulli.

Applications

Application 1: First order variance bounds

Family of first order lower and upper covariance bounds for functionals of arbitrary univariate distributions.

$\rightsquigarrow$ E., Reinert and Swan (2019a), Bernoulli.

Application 2: Infinite covariance expansions

Covariance expansions for functionals of arbitrary univariate distributions.

$\rightsquigarrow$ E., Reinert and Swan (2019b), arXiv:1906.08376

Applications

Application 1: First order variance bounds

Family of first order lower and upper covariance bounds for functionals of arbitrary univariate distributions.

$\rightsquigarrow$ E., Reinert and Swan (2019a), Bernoulli.

Application 2: Infinite covariance expansions

Covariance expansions for functionals of arbitrary univariate distributions.

$\rightsquigarrow$ E., Reinert and Swan (2019b), arXiv:1906.08376

Application 3: Stein factor and distances

Estimates of distances between univariate probability distributions.

$\rightsquigarrow$ E. and Swan (2019), arXiv:1909.11518

Application 1: first order variance bounds

Theorem (Cramer-Rao and weighted Poincaré inequality)

$$\frac{\mathbb{E}\left[-\mathcal{L}_p^\ell h(X)(\Delta^{-\ell} g(X))\right]^2}{\mathrm{Var}\big(h(X)\big)} \le \mathrm{Var}[g(X)] \le \mathbb{E}\left[(\Delta^{-\ell} g(X))^2 \frac{-\mathcal{L}_p^\ell h(X)}{\Delta^{-\ell} h(X)}\right]$$

for decreasing h with equality if and only if $g \propto h$.

Application 1: first order variance bounds

Theorem (Cramer-Rao and weighted Poincaré inequality)

$$\frac{\mathbb{E}\left[-\mathcal{L}_p^\ell h(X)(\Delta^{-\ell} g(X))\right]^2}{\mathrm{Var}\big(h(X)\big)} \leq \mathrm{Var}[g(X)] \leq \mathbb{E}\left[(\Delta^{-\ell} g(X))^2 \frac{-\mathcal{L}_p^\ell h(X)}{\Delta^{-\ell} h(X)}\right]$$

for decreasing h with equality if and only if $g \propto h$.

Examples

- (Chernoff, 1980) If $X \sim \mathcal{N}(0,1)$ then

$$\mathbb{E}[g'(X)]^2 \leq \mathrm{Var}[g(X)] \leq \mathbb{E}[(g'(X))^2]$$

- (Brascamp-Lieb, 1976) For appropriate density and g,

$$\frac{\mathbb{E}\left[(\Delta^{-\ell} g(X))\right]^2}{\mathbb{E}\left[\left(\rho_p^\ell(X)\right)^2\right]} \leq \mathrm{Var}[g(X)] \leq \mathbb{E}\left[\frac{(\Delta^{-\ell} g(X))^2}{-\Delta^{-\ell}\rho_p^\ell(X)}\right]$$

Application 2: infinite expansion

Application 2: infinite expansion

Papathanasiou-type expansion

$$\mathrm{Cov}\left[\mathbf{f}(X)\right] = \sum_{k=1}^{n}(-1)^{k-1}\mathbb{E}\left[\Delta^{-\ell_k}\mathbf{f}_{k-1}(X)\Delta^{-\ell_k}\mathbf{f}'_{k-1}(X)\frac{\Gamma_k^{\boldsymbol{\ell}}\mathbf{h}(X)}{\Delta^{-\ell_k}h_k(X)}\right] + (-1)^n R_n^{\boldsymbol{\ell}}(\mathbf{h})$$

for weight sequences $\Gamma_k^{\boldsymbol{\ell}}\mathbf{h}(x)$ and remainder term $R_n^{\boldsymbol{\ell}}(\mathbf{h})$.

Application 2: infinite expansion

Papathanasiou-type expansion

$$\mathrm{Cov}\left[\mathbf{f}(X)\right] = \sum_{k=1}^{n}(-1)^{k-1}\mathbb{E}\left[\Delta^{-\ell_k}\mathbf{f}_{k-1}(X)\Delta^{-\ell_k}\mathbf{f}'_{k-1}(X)\frac{\Gamma_k^{\boldsymbol{\ell}}\mathbf{h}(X)}{\Delta^{-\ell_k}h_k(X)}\right] + (-1)^n R_n^{\boldsymbol{\ell}}(\mathbf{h})$$

for weight sequences $\Gamma_k^{\boldsymbol{\ell}}\mathbf{h}(x)$ and remainder term $R_n^{\boldsymbol{\ell}}(\mathbf{h})$.

Application 2: infinite expansion

Papathanasiou-type expansion

$$\mathrm{Cov}\left[\mathbf{f}(X)\right] = \sum_{k=1}^{n}(-1)^{k-1}\mathbb{E}\left[\Delta^{-\ell_k}\mathbf{f}_{k-1}(X)\Delta^{-\ell_k}\mathbf{f}'_{k-1}(X)\frac{\Gamma_k^{\boldsymbol{\ell}}\mathbf{h}(X)}{\Delta^{-\ell_k}h_k(X)}\right] + (-1)^n R_n^{\boldsymbol{\ell}}(\mathbf{h})$$

for weight sequences $\Gamma_k^{\boldsymbol{\ell}}\mathbf{h}(x)$ and remainder term $R_n^{\boldsymbol{\ell}}(\mathbf{h})$.

References:

- $h_k(x) = x$: Papathanasiou (1988, $\ell_k = 0$)
- $h_k(x) = x$: Afendras et al. (2007, $\ell_k = -1$)
- $n = 1$ and $h_1 = -(\log p)'$ for log-concave p: Brascamp-Lieb inequality

Examples

Normal expansion

if $X \sim \mathcal{N}(0,1)$ then $\Gamma_k^0(x) = \frac{1}{k!}$ for all k and

$$\mathrm{Cov}[f(X), g(X)] = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k!} \mathbb{E}\left[f^{(k)}(X) g^{(k)}(X)\right].$$

Examples

Normal expansion

if $X \sim \mathcal{N}(0,1)$ then $\Gamma_k^0(x) = \frac{1}{k!}$ for all k and

$$\mathrm{Cov}[f(X), g(X)] = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k!} \mathbb{E}\left[f^{(k)}(X) g^{(k)}(X)\right].$$

Beta expansion

If $X \sim \mathrm{Beta}(a,b)$ then $\Gamma_k^0(x) = \frac{(x(1-x))^k}{k!(a+b)^{[k]}}$ for all k and

$$\mathrm{Cov}[f(X), g(X)] = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k!(a+b)^{[k]}} \mathbb{E}\left[f^{(k)}(X) g^{(k)}(X) X^k (1-X)^k\right].$$

Examples (2)

Poisson expansion

If $X \sim \mathrm{Poi}(\lambda)$, the order 1 expansions are

$$\begin{aligned}\mathrm{Var}[g(X)] &= \mathbb{E}\left[X(\Delta^- g(X))^2\right] - R_1^+ \\ &= \lambda\mathbb{E}\left[(\Delta^+ g(X))^2\right] - R_1^-;\end{aligned}$$

Examples (2)

Poisson expansion

If $X \sim \mathrm{Poi}(\lambda)$, the order 1 expansions are

$$\begin{aligned}\mathrm{Var}[g(X)] &= \mathbb{E}\left[X(\Delta^{-}g(X))^2\right] - R_1^{+} \\ &= \lambda\mathbb{E}\left[(\Delta^{+}g(X))^2\right] - R_1^{-};\end{aligned}$$

and the order 2 expansions are

$$\begin{aligned}\mathrm{Var}[g(X)] &= \mathbb{E}\left[X(\Delta^{-}g(X))^2\right] - \frac{1}{2}\mathbb{E}\left[X(X-1)(\Delta^{--}g(X))^2\right] - R_2^{++} \\ &= \mathbb{E}\left[X(\Delta^{-}g(X))^2\right] - \frac{1}{2}\lambda\mathbb{E}\left[X(\Delta^{-+}g(X))^2\right] - R_2^{+-} \\ &= \lambda\mathbb{E}\left[(\Delta^{+}g(X))^2\right] - \frac{1}{2}\lambda\mathbb{E}\left[X(\Delta^{+-}g(X))^2\right] - R_2^{-+} \\ &= \lambda\mathbb{E}\left[(\Delta^{+}g(X))^2\right] - \frac{1}{2}\lambda^2\mathbb{E}\left[(\Delta^{++}g(X))^2\right] - R_2^{--}\end{aligned}$$

Application 3: Stein factors and distances

Let X_n and X_∞ be random variables.

Application 3: Stein factors and distances

Let X_n and X_∞ be random variables.

Stein identity (IBP 1)

Two linear operators $\mathcal{T}_\infty^\ell$ and $\mathcal{L}_\infty^\ell$ are associated with X_∞ such that

$$\mathbb{E}\left[\left(\mathcal{T}_\infty^\ell c(X_\infty)\right) g(X_\infty)\right] = -\mathbb{E}\left[c(X_\infty)\Delta^{-\ell} g(X_\infty)\right]$$

are valid for all sufficiently regular functions c, g.

Application 3: Stein factors and distances

Let X_n and X_∞ be random variables.

Stein identity (IBP 1)

Two linear operators $\mathcal{T}_\infty^\ell$ and $\mathcal{L}_\infty^\ell$ are associated with X_∞ such that

$$\mathbb{E}\left[\left(\mathcal{T}_\infty^\ell c(X_\infty)\right) g(X_\infty)\right] = -\mathbb{E}\left[c(X_\infty)\Delta^{-\ell} g(X_\infty)\right]$$

are valid for all sufficiently regular functions c, g.

In particular, if $X_n \overset{\mathcal{L}}{\neq} X_\infty$, there exist some g functions such that

$$\mathbb{E}\left[\left(\mathcal{T}_\infty^\ell c(X_n)\right) g(X_n)\right] \neq -\mathbb{E}\left[c(X_n)\Delta^{-\ell} g(X_n)\right].$$

Application 3: Stein factors and distances

Let X_n and X_∞ be random variables.

Stein identity (IBP 1)

Two linear operators $\mathcal{T}^\ell_\infty$ and $\mathcal{L}^\ell_\infty$ are associated with X_∞ such that

$$\mathbb{E}\left[\left(\mathcal{T}^\ell_\infty c(X_\infty)\right)g(X_\infty)\right] = -\mathbb{E}\left[c(X_\infty)\Delta^{-\ell}g(X_\infty)\right]$$

are valid for all sufficiently regular functions c, g.

In particular, if $X_n \overset{\mathcal{L}}{\neq} X_\infty$, there exist some g functions such that

$$\mathbb{E}\left[\left(\mathcal{T}^\ell_\infty c(X_n)\right)g(X_n)\right] \neq -\mathbb{E}\left[c(X_n)\Delta^{-\ell}g(X_n)\right].$$

Stein discrepancy

Quantify the "distance" between the laws of X_n and X_∞:

$$\mathcal{S}(X_n, X_\infty, \mathcal{G}) := \sup_{g\in\mathcal{G}}\left|\mathbb{E}\left[\left(\mathcal{T}^\ell_\infty c(X_n)\right)g(X_n) + c(X_n)\Delta^{-\ell}g(X_n)\right]\right|$$

IPM and Stein discrepancy

Integral Probability Metric (IPM)

$$\mathcal{D}_{\mathcal{H}}(X_n, X_\infty) = \sup_{h \in \mathcal{H}} |\mathbb{E}h(X_n) - \mathbb{E}h(X_\infty)|$$

IPM and Stein discrepancy

Integral Probability Metric (IPM)

$$\mathcal{D}_{\mathcal{H}}(X_n, X_\infty) = \sup_{h \in \mathcal{H}} |\mathbb{E}h(X_n) - \mathbb{E}h(X_\infty)|$$

Examples

- Kolmogorov distance: $\sup_{z \in \mathbb{R}} |\mathbb{P}(X_n \le z) - \mathbb{P}(X_\infty \le z)|$
  is associated with $\mathcal{H}_{\mathrm{Kol}} = \{h(x) = \mathbb{I}[x \in (-\infty, z]] : z \in \mathbb{R}\}$.

IPM and Stein discrepancy

Integral Probability Metric (IPM)

$$\mathcal{D}_{\mathcal{H}}(X_n, X_\infty) = \sup_{h \in \mathcal{H}} |\mathbb{E}h(X_n) - \mathbb{E}h(X_\infty)|$$

Examples

- Kolmogorov distance: $\sup_{z \in \mathbb{R}} |\mathbb{P}(X_n \le z) - \mathbb{P}(X_\infty \le z)|$
  is associated with $\mathcal{H}_{\mathrm{Kol}} = \{h(x) = \mathbb{I}[x \in (-\infty, z]] : z \in \mathbb{R}\}$.
- Total variation distance: $\sup_{B \subset \mathbb{R}} |\mathbb{P}(X_n \in B) - \mathbb{P}(X_\infty \in B)|$
  is associated with $\mathcal{H}_{\mathrm{TV}} = \{h(x) = \mathbb{I}[x \in B] : B \in \mathcal{B}(\mathbb{R})\}$.

IPM and Stein discrepancy

Integral Probability Metric (IPM)

$$\mathcal{D}_{\mathcal{H}}(X_n, X_\infty) = \sup_{h \in \mathcal{H}} |\mathbb{E}h(X_n) - \mathbb{E}h(X_\infty)|$$

Examples

- Kolmogorov distance: $\sup_{z \in \mathbb{R}} |\mathbb{P}(X_n \leq z) - \mathbb{P}(X_\infty \leq z)|$ is associated with $\mathcal{H}_{\mathrm{Kol}} = \{h(x) = \mathbb{I}[x \in (-\infty, z]] : z \in \mathbb{R}\}$.
- Total variation distance: $\sup_{B \subset \mathbb{R}} |\mathbb{P}(X_n \in B) - \mathbb{P}(X_\infty \in B)|$ is associated with $\mathcal{H}_{\mathrm{TV}} = \{h(x) = \mathbb{I}[x \in B] : B \in \mathcal{B}(\mathbb{R})\}$.
- Wasserstein distance: $\int_{-\infty}^{\infty} |\mathbb{P}(X_n \leq z) - \mathbb{P}(X_\infty \leq z)| \, \mathrm{d}z$ is associated with $\mathcal{H}_{\mathrm{Wass}} = \{h : |h(x) - h(y)| \leq |x - y| \}$.

IPM and Stein discrepancy

Integral Probability Metric (IPM)

$$\mathcal{D}_{\mathcal{H}}(X_n, X_\infty) = \sup_{h \in \mathcal{H}} |\mathbb{E}h(X_n) - \mathbb{E}h(X_\infty)|$$

Theorem

$$\mathcal{D}_{\mathcal{H}}(X_n, X_\infty) = \mathcal{S}(X_n, X_\infty, \mathcal{G}_{\mathcal{H}})$$

where

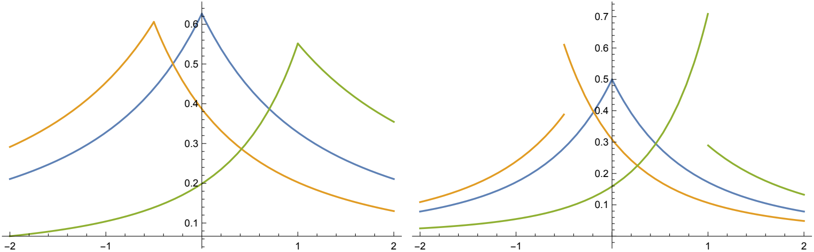
$$\mathcal{G}_{\mathcal{H}} = \left\{ g_h(x) = \frac{\mathcal{L}_p^\ell h(x+\ell)}{c(x+\ell)} : h \in \mathcal{H} \right\}.$$

Examples

Lower half-line indicator: $h(x) = \mathbb{I}[x \leq \xi]$ ($\ell = 0$)

- $g(x) = \frac{1}{c(x)} \frac{P(\xi \wedge x)\bar{P}(\xi \vee x)}{p(x)}$
- $g'(x) = \frac{\mathbb{I}[x \leq \xi] - P(\xi)}{c(x)} - \frac{\mathcal{T}_p^0 c(x)}{c^2(x)} \frac{P(\xi \wedge x)\bar{P}(\xi \vee x)}{p(x)}$.

$g(x)$ and $|g'(x)|$ for Gaussian target



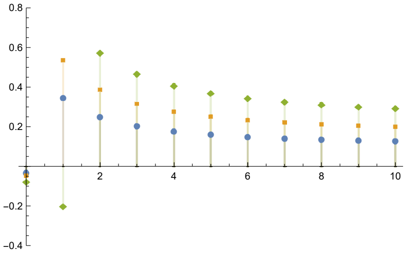
Examples

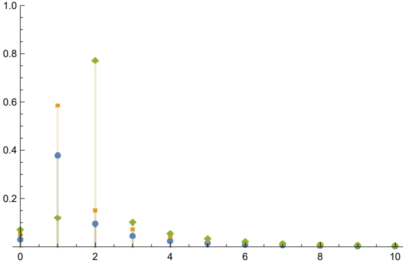
Point mass: $h(x) = \mathbb{I}[x = \xi]$ $(\ell = \pm 1)$

- $g(x) = \frac{p(\xi)}{c(x+\ell)p(x+\ell)} \left(\mathbb{I}[x \geq \xi + b_\ell] - P(x - b_\ell)\right)$
- If $c = \tau_p^\ell$,

$$\Delta^{-\ell} g_\xi^\ell(x) = \frac{\mathbb{I}[x{=}\xi] - p(\xi)}{\tau_p^+(x)} + \frac{p(\xi)(\mathbb{I}[x{\geq}\xi] - P(x))}{p(x)} \left(\frac{1}{\tau_p^-(x)} - \frac{1}{\tau_p^+(x)}\right)$$

$g(x)$ and $|\Delta^- g(x)|$ for Poisson target $(\ell = 1)$





Stein factor for general h

Proposition

Let $\kappa_1 = \sup_{y \in \mathcal{S}(p)} h(y) - \inf_{y \in \mathcal{S}(p)} h(y)$ and $\kappa_2 = \sup_{y \in \mathcal{S}(p)} |\Delta^{-\ell} h(y)|$.

Stein factor for general h

Proposition

Let $\kappa_1 = \sup_{y \in \mathcal{S}(p)} h(y) - \inf_{y \in \mathcal{S}(p)} h(y)$ and $\kappa_2 = \sup_{y \in \mathcal{S}(p)} |\Delta^{-\ell} h(y)|$.

1. If h is bounded then
   - $|g(x)| \leq \kappa_1 \frac{P(x-b_\ell)\bar{P}(x-b_\ell)}{p(x+\ell)} \frac{1}{c(x+\ell)}$
   - $\left|\Delta^{-\ell} g(x)\right| \leq \kappa_1 \frac{1}{c(x)} \left(1 + \frac{|\mathcal{T}_p^\ell c(x)|}{c(x+\ell)} \frac{P(x-b_\ell)\bar{P}(x-b_\ell)}{p(x+\ell)}\right).$

Stein factor for general h

Proposition

Let $\kappa_1 = \sup_{y \in \mathcal{S}(p)} h(y) - \inf_{y \in \mathcal{S}(p)} h(y)$ and $\kappa_2 = \sup_{y \in \mathcal{S}(p)} |\Delta^{-\ell} h(y)|$.

1. If h is bounded then
   - $|g(x)| \leq \kappa_1 \frac{P(x-b_\ell)\bar{P}(x-b_\ell)}{p(x+\ell)} \frac{1}{c(x+\ell)}$
   - $\left|\Delta^{-\ell} g(x)\right| \leq \kappa_1 \frac{1}{c(x)} \left(1 + \frac{|\mathcal{T}_p^\ell c(x)|}{c(x+\ell)} \frac{P(x-b_\ell)\bar{P}(x-b_\ell)}{p(x+\ell)}\right).$
2. If $\Delta^{-\ell} h$ exists and is bounded then
   - $|g(x)| \leq \kappa_2 \frac{\tau_p^\ell(x+\ell)}{c(x+\ell)}$
   - $\left|\Delta^{-\ell} g(x)\right| \leq \kappa_2 \left(\frac{|x-\mathbb{E}[X]|}{c(x)} + \frac{|\mathcal{T}_p^\ell c(x)|}{c(x)} \frac{\tau_p^\ell(x+\ell)}{c(x+\ell)}\right).$

Stein factor for general h

Proposition

Let $\kappa_1 = \sup_{y \in \mathcal{S}(p)} h(y) - \inf_{y \in \mathcal{S}(p)} h(y)$ and $\kappa_2 = \sup_{y \in \mathcal{S}(p)} |\Delta^{-\ell} h(y)|$.

1. If h is bounded then
   - $|g(x)| \leq \kappa_1 \frac{P(x-b_\ell)\bar{P}(x-b_\ell)}{p(x+\ell)} \frac{1}{c(x+\ell)}$
   - $\left|\Delta^{-\ell} g(x)\right| \leq \kappa_1 \frac{1}{c(x)} \left(1 + \frac{|\mathcal{T}_p^\ell c(x)|}{c(x+\ell)} \frac{P(x-b_\ell)\bar{P}(x-b_\ell)}{p(x+\ell)}\right).$
2. If $\Delta^{-\ell} h$ exists and is bounded then
   - $|g(x)| \leq \kappa_2 \frac{\tau_p^\ell(x+\ell)}{c(x+\ell)}$
   - $\left|\Delta^{-\ell} g(x)\right| \leq \kappa_2 \left(\frac{|x-\mathbb{E}[X]|}{c(x)} + \frac{|\mathcal{T}_p^\ell c(x)|}{c(x)} \frac{\tau_p^\ell(x+\ell)}{c(x+\ell)}\right).$

If, moreover, $c = -\mathcal{L}_p^\ell \eta$, three other bounds are provided.

Stein factor for general h

Proposition

Let $\kappa_1 = \sup_{y \in \mathcal{S}(p)} h(y) - \inf_{y \in \mathcal{S}(p)} h(y)$ and $\kappa_2 = \sup_{y \in \mathcal{S}(p)} |\Delta^{-\ell} h(y)|$.

1. If h is bounded then
   - $|g(x)| \leq \kappa_1 \frac{P(x-b_\ell)\bar{P}(x-b_\ell)}{p(x+\ell)} \frac{1}{c(x+\ell)}$
   - $\left|\Delta^{-\ell} g(x)\right| \leq \kappa_1 \frac{1}{c(x)} \left(1 + \frac{|\mathcal{T}_p^\ell c(x)|}{c(x+\ell)} \frac{P(x-b_\ell)\bar{P}(x-b_\ell)}{p(x+\ell)}\right).$
2. If $\Delta^{-\ell} h$ exists and is bounded then
   - $|g(x)| \leq \kappa_2 \frac{\tau_p^\ell(x+\ell)}{c(x+\ell)}$
   - $\left|\Delta^{-\ell} g(x)\right| \leq \kappa_2 \left(\frac{|x-\mathbb{E}[X]|}{c(x)} + \frac{|\mathcal{T}_p^\ell c(x)|}{c(x)} \frac{\tau_p^\ell(x+\ell)}{c(x+\ell)}\right).$

If, moreover, $c = -\mathcal{L}_p^\ell \eta$, three other bounds are provided.

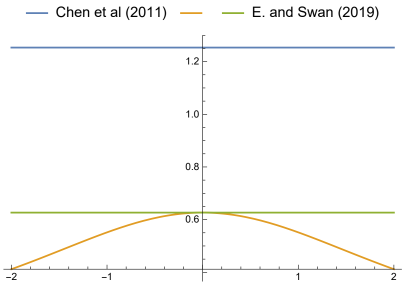
Some references for $\ell = 0$: Döbler (2012, 2015)

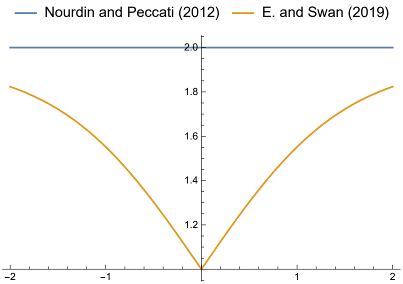
Application to specific distributions

Gaussian distribution

$$|g(x)| \leq \kappa_1 \frac{\Phi(x)(1-\Phi(x))}{\varphi(x)} \leq \kappa_1 \frac{1}{2}\sqrt{\frac{\pi}{2}}$$

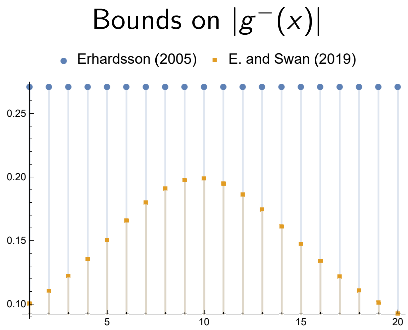
$$|g'(x)| \leq \kappa_1 \left(1 + |x| \frac{\Phi(x)(1-\Phi(x))}{\varphi(x)}\right) \leq 2\kappa_1$$

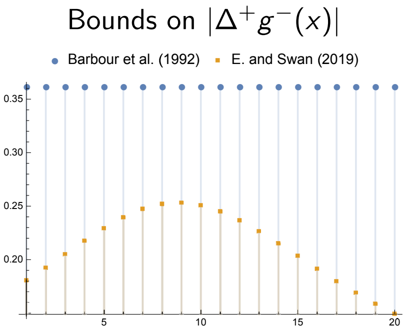




Application to specific distributions

Poisson distribution with parameter 10





Bounds on IPM's

Theorem: Stein discrepancies

Let $X_n \sim p_n$ be some random variable and let X_∞ have canonical Stein operators $\mathcal{T}_\infty^{\ell_\infty}$ and $\mathcal{L}_\infty^{\ell_\infty}$. Then,

$$\begin{aligned}
\mathbb{E}h(X_n) - \mathbb{E}h(X_\infty) &= \mathbb{E}\left[(\eta_1(X_n) - \mathbb{E}[\eta_1(X_\infty)])\frac{\mathcal{L}_\infty^{\ell_\infty} h(X_n + \ell_\infty)}{\mathcal{L}_\infty^{\ell_\infty}\eta(X_n + \ell_\infty)}\right] \\
&\quad + \mathbb{E}\left[\mathcal{L}_\infty^{\ell_\infty}\eta(X_n)\Delta^{-\ell_\infty}\left(\frac{\mathcal{L}_\infty^{\ell_\infty} h(\cdot + \ell_\infty)}{\mathcal{L}_\infty^{\ell_\infty}\eta(\cdot + \ell_\infty)}\right)(X_n)\right] \\
&= \mathbb{E}\left[\left(\mathcal{T}_\infty^{\ell_\infty} c_1(X_n)\right)\frac{\mathcal{L}_\infty^{\ell_\infty} h(X_n + \ell_\infty)}{c_1(X_n + \ell_\infty)}\right] \\
&\quad + \mathbb{E}\left[c_1(X_n)\Delta^{-\ell_\infty}\left(\frac{\mathcal{L}_\infty^{\ell_\infty} h(\cdot + \ell_\infty)}{c_1(\cdot + \ell_\infty)}\right)(X_n)\right].
\end{aligned}$$

In particular, IPM can be written as suprema of either of the above.

Example: Gaussian target

$X_\infty \sim \mathcal{N}(0,1)$

Total variation distance

$$\mathrm{TV}(X_n, X_\infty) \leq \frac{1}{2}\sqrt{\frac{\pi}{2}}\mathbb{E}\left[|X_n + \rho_n(X_n)|\right] + \sup_z \kappa_1^\star(z)$$

Example: Gaussian target

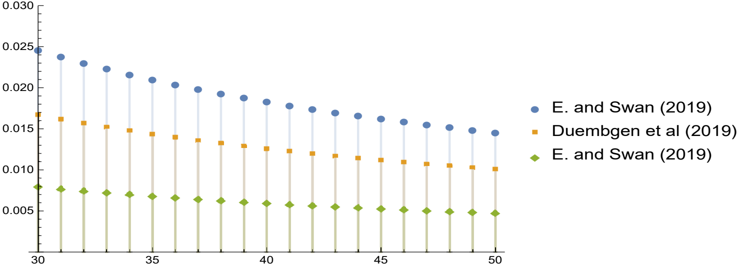
$X_\infty \sim \mathcal{N}(0,1)$

Total variation distance

$$\mathrm{TV}(X_n, X_\infty) \leq \frac{1}{2}\sqrt{\frac{\pi}{2}}\mathbb{E}\left[|X_n + \rho_n(X_n)|\right] + \sup_z \kappa_1^\star(z)$$

Normal vs Student ($X_n \sim t_n$)

$$\mathrm{TV}(X_n, X_\infty) \leq \mathbb{E}\left[|X_n|\left|\frac{X_n^2 - 1}{X_n^2 + n}\right|\frac{\Phi(X_n)\bar{\Phi}(X_n)}{\varphi(X_n)}\right] \leq \frac{2/\sqrt{e} - 1/2}{n-1}$$



Example: Gaussian target

$X_\infty \sim \mathcal{N}(0,1)$

Wasserstein distance

$$\mathrm{Wass}(X_n, X_\infty) \leq \mathbb{E}\left[|\rho_n(X_n)+X_n|\right] + \sup_{h\in\mathrm{Lip}(1)} |\kappa_1^\star(h)|$$

Example: Gaussian target

$X_\infty \sim \mathcal{N}(0,1)$

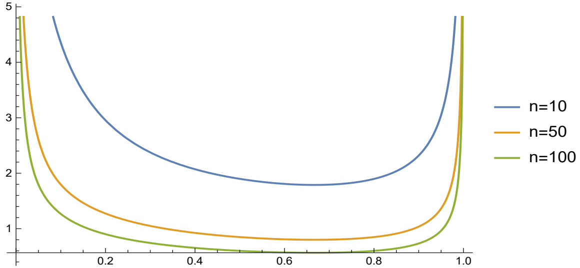
Wasserstein distance

$$\mathrm{Wass}(X_n, X_\infty) \le \mathbb{E}\left[|\rho_n(X_n) + X_n|\right] + \sup_{h \in \mathrm{Lip}(1)} |\kappa_1^\star(h)|$$

Normal vs Binomial

If X_n is standardized binomial with parameters (n, θ),

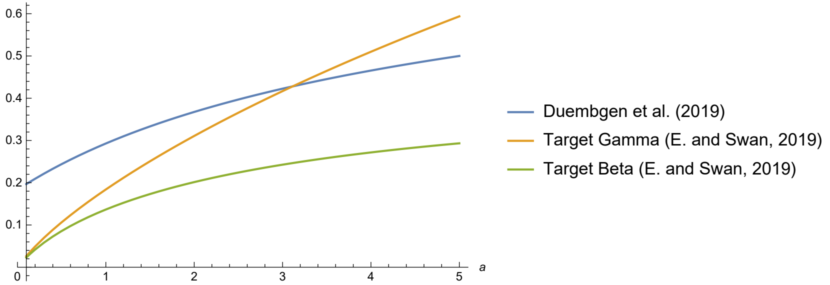
$$\mathrm{Wass}(X_n, X_\infty) \le 2\sqrt{\frac{1}{\theta} - 1}\frac{1}{\sqrt{n}} + \frac{2}{\sqrt{n\theta(1-\theta)}} + (1-\theta)^n$$



Example: Beta vs Gamma

TV between Beta$(a, 3)$ and Gamma$(a, a+3)$

We can choose $X_\infty \sim$ Beta$(a, 3)$ and $X_n \sim$ Gamma$(a, a+3)$ or the opposite.



Conclusion

Part 1: spatial data

- Ernst, M. and G. Haesbroeck (2017). Comparison of local outlier detection techniques in spatial multivariate data. *Data Mining and Knowledge Discovery* 31(2), 371–399.
- Ernst, M. and G. Haesbroeck. Robustness of tests for spatial autocorrelation.

Conclusion

Part 1: spatial data

- Ernst, M. and G. Haesbroeck (2017). Comparison of local outlier detection techniques in spatial multivariate data. *Data Mining and Knowledge Discovery* 31(2), 371–399.
- Ernst, M. and G. Haesbroeck. Robustness of tests for spatial autocorrelation.

Part 2: Stein's method

- Ernst, M., G. Reinert, and Y. Swan (2019). First order covariance inequalities via Stein's method. *Bernoulli*. In press.
- Ernst, M., G. Reinert, and Y. Swan (2019). On infinite covariance expansions. arXiv:1906.08376.
- Ernst, M. and Y. Swan (2019). Distances between distributions via Stein's method. arXiv:1909.11518.

Thank you

Moran
estimator
panels
weight
attributes
assumptions
distance
covariance
resistance
Table
Proof
contamination
distances
matrix
grid
mean
tests
random
detection
measure
value
instance
Probability
Harris
multivariate
Section
distribution
weights
method
operators
functions
local
obtain
given
results
Stein
spatial
data
global
technique
function
follows
lower
sample
test
set
may
class
operator
simulations
inequality
continuous
outlier
variance
techniques
Poisson
consider
holds
respect
location
statistics